Genomics of invasion: Diversity and selection in introduced populations of monkeyflowers (*Mimulus guttatus*)

Joshua Puzey[1,2]

Mario Vallejo-Marín[3]

[1] Department of Biology, Duke University, Durham, North Carolina, United States of America. 27705.

[2] Department of Biology, College of William and Mary, Williamsburg, Virginia, United States of America. 23185.

[3] Biological and Environmental Sciences, University of Stirling. Stirling, Scotland, United Kingdom. FK9 4LA. Email: mario.vallejo@stir.ac.uk

Keywords: anthropogenic dispersal; genome scan; introduced species; long-distance colonisation; next generation sequencing; selective sweeps.

Running title: Genomic analysis of introduced monkeyflowers

ABSTRACT

Global trade and travel is irreversibly changing the distribution of species around the world. Because introduced species experience drastic demographic events during colonisation, and often face novel environmental challenges from their native range, introduced populations may undergo rapid evolutionary change. Genomic studies provide the opportunity to investigate the extent to which demographic, historical, and selective processes shape the genomic structure of introduced populations by analysing the signature that these processes leave on genomic variation. Here we use next-generation sequencing to compare genome-wide relationships and patterns of diversity in native and introduced populations of the yellow monkeyflower (*Mimulus guttatus*). Genome resequencing data from ten introduced populations from the United Kingdom (UK) and 12 native *M. guttatus* populations in North America (NA), demonstrated reduced neutral genetic diversity in the introduced range, and showed that UK populations are derived from a geographic region around the North Pacific. A selective-sweep analysis revealed site frequency changes consistent with selection on 5 of 14 chromosomes, with genes in these regions showing reduced silent site diversity. While the target of selection is unknown, genes associated with flowering time and biotic and abiotic stresses were identified within the swept regions. The future identification of the specific source of origin of introduced UK populations will help determining if the observed selective sweeps can be traced to un-sampled native populations or occurred since dispersal across the Atlantic. Our study demonstrates the general potential of genome-wide analyses to uncover a range of evolutionary processes affecting invasive populations.

INTRODUCTION

The introduction of species beyond their native ranges can affect ecological and evolutionary interactions in the new habitat (Cox 2004; Phillips & Shine 2006; Liu & Pemberton 2009; Ricciardi *et al.* 2013), and can negatively impact levels of local biodiversity, and result in high economic costs (Pimentel 2002; Williams *et al.* 2010; Vila *et al.* 2011). Introduced populations are often used models to investigate rapid genetic changes and adaptation to novel environments, thus providing valuable insights into basic biological processes including local adaptation (Sax *et al.* 2007; Prentis *et al.* 2008). In particular, genetic analyses continue to play a central role in studies of the origin and establishment of introduced populations, as well as of the mechanisms that permit the colonisation and drive the spread of populations beyond their native range (Baker & Stebbins 1965; Lee 2002).

The genomic structure of non-native populations is influenced by a variety of processes including population bottlenecks, multiple introductions, population expansion, gene flow between populations, and selection, among others (Lee 2002). For instance, in populations established after limited long distance dispersal events, the level of genetic diversity can be significantly lower than in the native range, reflecting population bottlenecks (Lachmuth *et al.* 2010; Ness *et al.* 2012). However, introduction of multiple individuals from the same population, or multiple introductions from genetically diverse source populations, can counteract the loss of diversity or even result in higher levels of genetic variation within introduced populations compared to native ones (Dlugosch & Parker 2008). The level of standing variation in introduced populations is relevant to the colonisation process, as severe

bottlenecks and reduced diversity could indicate potential limitations for the rapid evolution of adaptive traits in novel environments (Barrett & Schluter 2008; Lachmuth *et al.* 2010; Siol *et al.* 2010; Messer & Petrov 2013). Severe bottlenecks resulting in globally reduced diversity may indicate that natural selection is mutation limited. However, genetic variation resulting from introduction of multiple individuals can provide ample standing variation for natural selection.

Genome wide studies have been employed to investigate genetic patterns in natural populations, including the relationship between native and introduced populations as well as invasion pathways of exotic plants and animals (Jahodová *et al.* 2007; Dlugosch *et al.* 2013; Tarnowska *et al.* 2013). Genome scans allow detecting selection acting on specific locations in the genome (Nielsen *et al.* 2005), and by comparing the sites under selection in the genomes of different populations, it is possible to identify candidates for genetic regions associated with local adaptation (Savolainen *et al.* 2013). A prerequisite to any genome wide study is identifying a large numbers of genetic markers, such as restriction site polymorphisms (e.g., AFLPs, Vos *et al.* 1995), or single nucleotide polymorphisms (SNPs). The growing access to high-throughput sequencing technologies at low costs opens the opportunity to conduct genome wide studies at an unprecedented depth, even in non-model organisms (Prentis *et al.* 2010; Twyford & Ennos 2012; Ellegren 2014).

The generation of genome wide markers by high throughput sequencing can employ methods for genome complexity reduction such as transcriptome sequencing (Dlugosch *et al.* 2013) or RAD sequencing (Davey *et al.* 2011; Roda *et al.* 2013). However, for the increasing number of species in which a reference genome is available, whole genome re-sequencing

allows genotyping markers, such as SNPs, which may occur at high densities across the genome (Davey *et al.* 2011; Twyford & Ennos 2012; Savolainen *et al.* 2013). Importantly, whole genome resequencing removes many of the ascertainment biases associated with SNP chips or other genome reduction technologies. The dense marker saturation achieved through genome re-sequencing is particularly useful for detecting the footprint of selection acting on specific locations across the genome. For instance, selective sweeps, in which selection drives previously rare alleles to fixation, also reduces diversity at neighbouring regions around the selected site (Messer & Petrov 2013). The signal left behind by selective sweeps can be detected by comparing patterns of variation along the genome with the level expected under a null model. Hard selective sweeps, where a single variant is driven to fixation, leave a characteristic footprint in the genome, which can be identified using summary statistics such Tajima's *D* or the composite likelihood ratio (CLR) (Nielsen *et al.* 2005; Messer & Petrov 2013). These statistics may be particularly powerful to detect recent selective sweeps as linkage disequilibrium (LD) between the selected site and the surrounding variation is expected to be highest immediately following the fixation of the adaptive allele.

Genomic studies of native and introduced populations can uncover demographic, historical, and selective processes by analysing the signature that these processes leave on genomic variation. Here we use whole genome re-sequencing to assess the relationship between native and introduced populations, and to uncover selective episodes in specific regions of the genome of introduced populations. We study the yellow monkeyflower (*Mimulus guttatus,* Phrymaceae), a species that has long been used as a model for ecological and evolutionary

104 studies in its native range (Vickery 1959; Wu *et al.* 2008), and which has become naturalised in
105 Eastern North American, New Zealand, Iceland, the Faroe Islands, and Western Europe (van
106 Kleunen & Fischer 2008; Murren *et al.* 2009; Tokarska-Guzik & Dajdok 2010), becoming
107 particularly widespread in the United Kingdom (Vallejo-Marín & Lye 2013). *Mimulus guttatus* is
108 ideally suited for studying the ecological genomics of non-native populations due to its recent
109 introduction and spread (<200 years), abundant information on the ecology and evolution of
110 native populations, and the availability of a full genome sequence, which provides a backbone
111 for analysing and interpreting patterns of genetic variation in introduced populations. The
112 relatively small genome of *M. guttatus* (1N = 430 MB), makes this species a good candidate for
113 population genomic studies through re-sequencing, as multiple individuals can be analysed with
114 a relatively small budget. We analysed previously available and newly generated whole genome
115 sequence data for 12 native and 10 introduced British populations of *M. guttatus*, as well as five
116 additional related taxa (*n* = 35 *Mimulus* genomes in total). Our data set allowed us to address
117 three specific aims: (1) to determine the level of genome-wide diversity present in introduced
118 populations of *M. guttatus* in the United Kingdom; (2) to investigate the genetic relationships of
119 native and introduced populations; and (3) to search for evidence of hard selective sweeps in
120 introduced populations.

METHODS

Study system

The *Mimulus guttatus* species complex includes a set of phenotypically variable, interfertile taxa with a native range of distribution in Western North America from northern Mexico to Alaska (Grant 1924; Wu *et al.* 2007). Within this complex, populations of *M. guttatus* Fischer ex DC. (Grant 1924) show marked variation in characteristics including life history (annual/perennial) (Hall & Willis 2006; Lowry & Willis 2010), mating system (Ritland 1990; Dole 1992), phenology (Hall & Willis 2006; Friedman & Willis 2013), floral morphology (Fishman *et al.* 2002), edaphic adaptations (e.g. tolerance to elevated concentrations of heavy metals or salt, Macnair & Watkins 1983; Lowry *et al.* 2008; Lowry *et al.* 2009), habitat preferences (Wu *et al.* 2008), chromosome number (most populations are diploid: 2n = 2x = 28, but tetraploids also occur in the native range, Sweigart *et al.* 2008), and clonal growth (Dole 1992; van Kleunen 2007), among others. This incredible diversity has led some taxonomists to subdivide *M. guttatus* into numerous morphological species (e.g., Pennell 1951; Nesom 2012). Here we adhere to the broader circumscription of *M. guttatus* Fischer ex DC. (Grant 1924; Wu *et al.* 2008).

Mimulus guttatus was introduced into the British Isles in 1812, and the first naturalised populations were reported in England around 1830 (Roberts 1964; Parker 1975). In the UK, *M. guttatus* is currently widespread and occurs in wet habitats along the banks of rivers and streams, in ditches, marshy areas, and other wet places (Stace 2010; Vallejo-Marín & Lye 2013). It propagates via both seeds and clonally through lateral stems that root freely at the nodes.

141 The source of the first naturalised populations of *M. guttatus* in the UK is unknown, but one of
142 the earliest specimens of this taxon to reach Europe was derived from material collected by
143 Langsdorff between 1806 and 1810 in the Aleutian Islands in Alaska, and transmitted to the
144 Botanic Gardens at Cambridge (Sims 1812; Pennell 1935, p. 116). The use of *Mimulus spp.* as a
145 horticultural species in Victorian England, as reflected by being readily available in seed
146 catalogues of the time (e.g., Gardeners' Chronicle 1852), raises the distinct possibility that *M.*
147 *guttatus* was introduced into the UK on repeated occasions and from multiple sources.

148 ## Population sampling

149 Analysing genomes across a wide geographic scale represents a trade-off between the numbers
150 of individuals *vs*. populations sampled. The goal of this study was to determine the introduction
151 history of *Mimulus* into the UK, the effects of the introduction on nucleotide diversity, and to
152 identify signals of selective sweeps that are common across the UK. To do this, we sought to
153 obtain samples from geographic disparate regions from across the UK. Obtaining geographically
154 distant samples increases the likelihood of identifying introductions from multiple different
155 donor populations. This sampling strategy also facilitated our goal of identifying selective
156 sweeps shared across the UK *M. guttatus* populations. Population specific selective sweeps
157 caused by local adaptation to narrow geographic and ecological niches in the UK are not
158 detected in our analyses and would require multiple individuals from the specific population of
159 interest. Previous molecular analyses of *Mimulus guttatus* have demonstrated that a scattered
160 sampling design, with 1 individual per population, is sufficient to capture regional
161 differentiation, and can avoid clustering biases resulting from sampling multiple individuals from

fewer populations (Oneal *et al.* 2014). In total, we analysed genome data from 27 populations: 12 *M. guttatus* populations in the native range, 10 UK *M. guttatus*, and five outgroups. From one of the native populations (Iron Mountain, IM) we sampled an additional 8 individuals, which allowed us to explore the sensitivity of our findings to the particular individual sampled within a population.

Introduced populations

We sampled 10 populations of *M. guttatus* spanning the range of distribution of this species in the British Isles (Table 1; Fig. 1). The northernmost population came from the Shetland Islands (QUA, N 60.105° W 1.227°) and the southernmost from Cornwall, England (CRO, N 50.163°, W 5.293°). A population from Northern Ireland was also included (VIC, N 54.763°, W 7.454°). Non-native populations were collected from banks of canals, streams or rivers (HOU, CER, VIC, AYR, DBL, TOM, and PAC), on roadside ditches (QUA), on waterlogged ground in an abandoned field (CRO), or in a bog near a small stream (TRE). A single wild-collected individual per population was randomly selected from each population for sequencing.

Native range populations and outgroups

We obtained sequence data from the Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra) from 12 native populations of *M. guttatus*, and five outgroups within section Simiolus: *M. nasutus* (SF), *M. cupriphillus* (MCN), *M. platycalyx* (CVP), *M. micranthus* (EBR), and *M. dentilobus* (DENT). The 12 native populations of *M. guttatus* covered a linear transect of approximately 2800km from Haida Gwaii (Queen Charlotte Islands), British Columbia (TSG, N 53.419°, W 131.916°) to Arizona (PED, N 32.711°, W 110.628°) (Table

1). Populations in the native range occurred in a diversity of wet habitats including river and stream banks, seeps, beach dunes, bogs, and springs. A single individual represented all but one of the native populations. In the case of the Iron Mountain population (IM), we were able to obtain data for nine separate individuals. Mean coverage per genotyped base per individual ranged between 4 – 29x with an average of 10x.

DNA isolation and sequencing

We collected leaf tissue of British *M. guttatus* individuals (one per population) in the field and preserved it in re-sealable plastic bags with self-indicating silica gel (Fisher Scientific, Loughborough, UK) at room temperature. This dry tissue was used for DNA extraction using the Leaf MasterPure total DNA extraction kit (Cambio Ltd, Cambridge, UK). DNA libraries were created and barcoded using the Nextera DNA sample preparation kit (Illumina, San Diego, California), which uses a transposon based method to randomly tag DNA for multiplexed sequencing. After library construction, an Agilent Bioanalyzer (Santa Clara, California) was used to measure length distribution of library, and a fluorometer (Qubit 2.0, Life Technologies, Paisley, UK) was used to measure concentration. Equimolar quantities of each library were pooled and sequenced in an Illumina HiSeq 2500 rapid-run producing 150 base-pair paired-end reads. Overall, we obtained raw coverage of 1.5 – 11x per individual with an average of 5.7x. Raw sequence data for UK *Mimulus* samples is deposited in the JGI SRA (SRA accession numbers will be available upon acceptance of the manuscript).

Sequence data analysis

Genome alignment and SNP genotyping

Raw reads were aligned to the *Mimulus guttatus* v2.0 genome available from Phytozome (http://www.phytozome.net) using *bowtie2* (Langmead & Salzberg 2014) using –fast-local allowing soft-clipping of poorly mapped read ends. After alignment, *Picard tools* (http://picard.sourceforge.net) was used to remove duplicates, add read groups, and verify that all mate information was accurate. After processing in *Picard tools*, the *Genome Analysis Toolkit* (GATK, DePristo *et al.* 2011) was used to call genotypes using the 'Unified Genotyper'. Minimum alignment quality was 25 and base quality was 25. Called genotypes were filtered to include genotypes with a call quality threshold of Q30 or greater. Insertions, deletions, and heterozygous sites were not included in subsequent analyses. Detailed command-line methods can be found in the Supplementary Information. After all filtering, mean coverage per genotyped base per individual for the 10 UK samples ranged from 1.7 – 5.8x with an average of 3.5x . Of the 293 Mb located on the main 14 genomic scaffolds (representing 14 linkage groups), after all filtering, 71 Mb were genotyped in at least one of the UK individuals. A total of 18.3, 18.5, and 8.9 Mb were genotyped in 8, 9, or 10 of the UK samples, respectively (Fig. S7).

Measures of genetic diversity

Nucleotide diversity at silent and non-silent sites was calculated using software described in Zhang *et al.* (2006). Briefly, genomes for all sequenced lines were recalled using the genotype data. Missing data was not imputed. Measurements of pairwise synonymous (π_{syn}) and non-

synonymous nucleotide diversity ($\pi_{non\text{-}syn}$) were calculated through pairwise comparison of coding sequences. Coding sequences were extracted from the recalled genomes using the gff3 gene annotation file available on phytozome.net. A Fisher p-value associated with each diversity value and indicates the confidence of that particular value. We only considered π_{syn} and $\pi_{non\text{-}yn}$ values for genes with a Fisher p-value ≤ 0.001 and alignment length greater than 200 bases. In addition to calculating nucleotide diversity at synonymous and non-synonymous sites, whole genome alignments were used to calculate genome-wide nucleotide diversity (π) in sliding windows using *VariScan* (Hutter *et al.* 2006). Windows of 50,000 genotyped bases and overlapping steps of 1000 bases were used.

Genetic relationships between introduced and native *M. guttatus*

In order to determine the genetic relationships between introduced and native populations we conducted an analysis of genetic similarity using a random subset of 1,400,000 SNPs. To create this data set, we randomly selected 100,000 SNPs for each of the 14 major linkage groups (chromosomes) that were genotyped in at least 30 individuals (out of 35). Our SNP data set is therefore not subject to ascertainment bias arising from selecting, for example, only coding or non-coding SNPs (Garvin *et al.* 2010). Instead the SNP dataset analysed here should represent a snapshot of the total genetic diversity of each sample and be shaped by both neutral and non-neutral processes (Helyar *et al.* 2011). Within each linkage group, neighbouring SNPs were separated by 209 bp on average (209 ± 3.34; mean ± SE). Each SNP was coded as "0" if it matched the reference allele, and "1" for the alternative allele. In this analysis we included all native and introduced individuals, and the five outgroups (*n* = 35 individuals). Multiple

243 individuals from IM were included as a reference of the variation seen within a single
244 population.

245 We constructed a genetic distance matrix using *p*-distance (the proportion of nucleotide
246 sites that differ between a pair of sequences) from the binary SNP data using the package *ape*
247 (Paradis *et al.* 2004) in *R* ver. 3.0.3 (R Development Core Team 2014). The combined distance
248 matrix was then used to estimate the relationships between all samples using a neighbour
249 joining (NJ) analysis in *ape*. Support for nodes in the NJ tree was calculated using 100 bootstrap
250 replicates. Trees were drawn using *FigTree* v. 1.4.0 (Rambaut 2014). The NJ distance-based
251 approach used here is appropriate for genome-wide analyses (e.g. Brandvain *et al.* 2014), as
252 maximum likelihood and Bayesian phylogenetic methods depend on specifying a mutational
253 model, which is not practical for genome-wide data. We also conducted a Principal Component
254 Analysis (PCA) using the function *glPca* in *adegenet* (Jombart & Ahmed 2011). This analysis
255 provides an independent estimate of the relationships between native and introduced
256 populations, and can be used to compare with the results of the NJ analysis. For the PCA, we
257 selected only one individual for each of the 12 native and 10 introduced populations of *M.*
258 *guttatus*. The identity of the particular individual chosen from the IM population had no
259 qualitative effect on the relationships inferred from the PCA (Fig S8), and similarly, randomly
260 choosing one IM individual instead of nine for the NJ analysis did not change the tree topology
261 (data not shown).

Selective sweep analysis

Regions in the genome showing the signature of selective sweeps were detected using the parametric approach described in Nielsen *et al.* (2005), and implemented in the program *SweepFinder*. This method compares the observed site frequency spectrum within local regions in the genome (windows) against the background site frequency spectrum seen across the entire genome (or linkage group), and calculates statistical departures from this background expectation using a composite likelihood ratio (CLR). Importantly, the null hypothesis employed by this method is derived from the background data itself and does not depend on specific population genetic models or assumptions about demographic equilibrium (Nielsen *et al.* 2005), which are unlikely to hold in recently introduced populations. *SweepFinder* is robust to models that include population growth with recombination (Nielsen et al. 2005).

One potential issue with *SweepFinder* is that it is sensitive to SNP density (Nielsen *et al.* 2005). To account for both shared ancestral sweeps and artefacts due to genotype density, we independently analysed the North American (NA) and UK data using the exact same criteria. Ten samples from the North American populations (AHQ, BOG, DUN, IM, LMC, MAR, PED, SWB, TSG, YJS) were chosen based on the results of the genome-wide relationship analysis. Next, independently, for both the NA and UK datasets, we determined whether a given site was polymorphic and asked how many individuals were genotyped at that particular site. Using this information, we choose sites that were genotyped in at least 8 individuals (of the 10 total) in both the UK and NA samples and were polymorphic in at least one of these populations. Thus, we ended up with a dataset that included the exact same number of sites from the exact same

283 genomic locations. Next, we calculated genome-wide SFS for each dataset (UK and NA). Each
284 chromosome was divided into 5000 bins and the SFS within each bin was compared to the
285 genome-wide SFS to look for signals of a selective sweep using the parametric approach
286 described (Nielsen *et al.* 2005) and implemented in *SweepFinder.*

287 After running *SweepFinder*, we independently plotted the genome-wide CLR distribution
288 for UK and NA samples. Within the NA samples, all genomic locations with CLR scores above the
289 median value were marked for masking. Liberal masking based on the NA analyses removes
290 hard sweeps that occurred in the last common ancestor, and removes artefacts due to variable
291 genotype density. The NA *SweepFinder* results were used to mask the UK genome *SweepFinder*
292 results. Only genomic positions that survived masking were considered in further analyses.
293 Within the UK, the top 1% CLR outliers were identified as regions that have possibly experienced
294 a hard sweep and subjected to further investigation. Gene coordinates are available in the gff3
295 gene annotation on Phytozome and genes with positions at least partially overlapping the swept
296 regions were extracted for further analyses.

297 # RESULTS

298 ## Nucleotide diversity in *M. guttatus*

299 Overall genome-wide nucleotide diversity in the UK was $\pi = 0.015$ (Fig. S1). Figures S2 and S3
300 show patterns of nucleotide diversity across the genome for both native and introduced
301 populations. For genes, within UK samples, diversity at silent sites was $\pi_{syn} = 0.0325$ while non-
302 synonymous diversity was $\pi_{non\text{-}syn} = 0.0035$ (Fig. S4). Within the native North American

303 populations, nucleotide diversity was calculated through comparisons of 10 individuals (same 10
304 individuals used as the NA samples in sweep analyses). Genome-wide diversity within the NA
305 samples is π= 0.031 (Fig. S1). Synonymous diversity within NA is π_{syn} = 0.0610 while non-
306 synonymous diversity is $\pi_{non\text{-}syn}$ = 0.0075 (Fig. S4). Comparing NA and UK nucleotide diversity
307 indicates an overall reduction in the introduced populations of approximately 50%.

308 **Genetic relationships of native and introduced populations**

309 The relationships between 22 native and introduced populations of *Mimulus guttatus*, and five
310 outgroups based on the genetic distance of 1,400,000 SNPs distributed across the genome is
311 shown in Fig. 2. All the population level nodes in this NJ tree had a bootstrap support of 100%.
312 The NJ tree shows that all 10 UK populations form a single well-supported clade (Fig. 2). The UK
313 clade is most closely related to the native TSG population, a coastal perennial *M. guttatus* from
314 Graham Island (Queen Charlotte Islands) in British Columbia (Lowry & Willis 2010). The UK and
315 TSG samples form part of a clade of populations located north of the N 40° parallel, and which
316 includes other inland perennial (BOG and YJS), and annual populations (AHQ, IM, and MAR)
317 (Figs. 1 and 2; Table 1). The NJ tree shows a second clade composed mainly of more southern
318 populations, and which includes inland plants (LMC, REM), coastal perennials (SWB, DUN), as
319 well as two annual outgroups (CVP, *M. platycalyx*; EBR, *M. micranthus*). The DUN population is
320 the only one in this group located north of the N 40° parallel (Fig. 1). Finally, the NJ tree shows a
321 third clade including three of outgroups (MCN, *M. cupriphilus;* SF, *M. nasutus;* DENT, *M.*
322 *dentilobus*), as well as the two most southern populations of *M. guttatus*, MED, an annual
323 inland population, and PED, an inland perennial from Arizona. Our results indicate that native

M. guttatus is separated into two main clades corresponding mostly to geographic location (North and South groups), and not to different life histories (annual/perennial) or habitat types (coastal/inland), as has been recently described by Brandvain *et al.* (2014). Finally, our NJ analysis also indicates that *M. cupriphilus* and *M. platycalyx* are nested within broadly circumscribed *M. guttatus* (Fig. 2).

The results of the PCA show clear support for a close relationship between all UK samples, and also indicate that the most genetically similar native population sampled here is the coastal perennial TSG (Fig. 3). The first principal component separates the North and South groups of native *M. guttatus*, with DUN partly overlapping with the North group. The second principal component separates the UK samples from most of the other northern accessions (Fig. 3). Together, the NJ and PCA show a common ancestry of British *Mimulus guttatus*, and its association with northern, native populations.

Evidence of selection in introduced populations

Our analysis of selective sweeps in native and introduced populations identified several genomic regions displaying changes in the site frequency spectrum (measured using the CLR statistic), consistent with the signatures of positive selection acting in these regions (Fig. 4; results for all linkage groups are shown in Fig. S5). The comparison of high CLR regions in the separate analyses conducted in NA and UK samples, allowed us to detect selective sweeps shared by both native and introduced populations. By masking these high CLR sites in NA, we located genomic regions that are candidates for selective sweeps occurring after the separation of the clade leading to the UK populations, including potentially unsampled North American donor

345 populations. Moreover, the masking also accounted for possible regions with high CLR scores
346 that were simply due to low genotyped SNP density. Our analysis revealed selective sweeps on
347 five of the 14 linkage groups in the UK, which are not shared with NA samples (Fig. 4, Table 1).
348 Regions with high CLR scores showed significantly reduced overall diversity: Diversity within
349 sweeps was $\pi = 0.0076$, half that of nucleotide diversity outside sweeps $\pi = 0.0152$. ($p \leq 0.0001$;
350 Figs. S6).

351 We identified a total of 299 genes located under the candidate region for selective
352 sweeps within the UK clade (Table S1). Synonymous diversity for genes within sweeps was
353 significantly lower than silent site diversity outside sweeps of ($\pi_{syn} = 0.0147$ vs. $\pi_{syn} = 0.0323$; $p \leq$
354 0.0001) (Fig. S6). Within sweeps, 28 genes had $\pi_{syn} < 0.01$ (Table S2) and only two genes under
355 sweeps had silent site diversity (π_{syn}) above the genome-wide mean. Taking advantage of the
356 annotated genome of *M. guttatus*, we recorded genes located within the swept regions, which
357 included genes involved in flowering time, abiotic stress response including nutrient transport
358 and freezer tolerance, and biotic stress responses (Table S1).

359 # DISCUSSION

360 Our study represents the first genome resequencing study of native and introduced populations
361 of *Mimulus guttatus.* By analysing whole genome sequences of 35 individuals from 22
362 populations we demonstrated that introduced plants in the UK are characterized by a ~50%
363 reduction in synonymous (π_{syn}) genetic diversity, and that UK populations form a single clade,
364 relative to the North American samples included here, suggesting a common origin for non-

native populations. Our analysis revealed changes in the site frequency spectrum at multiple locations across the genome, consistent with selective sweeps, some of which were restricted to UK samples. The reduction of both synonymous and non-synonymous nucleotide diversity in genes located within sweep regions, compared to genes outside of these sweeps, was consistent with the expected loss of genetic diversity in regions linked to selected loci. Future studies are required to assess whether selection is indeed responsible for the observed patterns of variation in the sweep candidate regions, as well as to determine if such selection acted before or after the introduction of *M. guttatus* into the UK. This study illustrates the potential of whole-genome sequencing studies to provide the initial steps for understanding the genomic consequences of invasion and rapid adaptation to new environments.

Origin and diversity of *Mimulus guttatus* populations in the UK

Determining the geographic origin of introduced populations provides a reference point for studies of the potential ecological and evolutionary changes occurring during the colonization and establishment phases of biological invasions (Milne & Abbott 2000). Our sample of North American populations includes a large part of the native range of *M. guttatus* (Fig. 1), but admittedly still represents a small fraction of populations from this widely distributed taxon (Grant 1924). However, we were able to include populations with different life histories, morphologies, and habitat preferences (Table 1), including different ecological and morphological groups (Lowry *et al.* 2008; Lowry & Willis 2010).

Our analysis of genome-wide polymorphism revealed two main clades of *M. guttatus* in the native range, which broadly correspond to their geographic origin (North and South groups;

386 Brandvain *et al.* 2014). An exception to the geographic arrangement of these two groups of *M.*
387 *guttatus* is the DUN population. Although geographically located in the northern range (Fig. 1),
388 DUN is nested within the southern group (Fig. 2), and may represent a secondary dispersal or
389 reflect an introgression event between these two groups (Brandvain *et al.* 2014). Interestingly,
390 the North and South clades of *M. guttatus* include populations with contrasting habitats,
391 morphologies, and life histories (Table 1), suggesting that taxonomic groupings based on
392 general morphological and ecological attributes are unlikely to correspond to monophyletic
393 clades (Nesom 2012). At the same time, the polyphyletic nature of the *M. guttatus* species
394 complex is reflected in our results by the fact that a *M. guttatus* population (MED) is nested
395 within outgroup taxa (SF: *M. nasutus*, and MCN: *M. cupriphilus;* Fig. 2), while *M. platycalyx* (CVP)
396 and *M. micranthus* (EBR) fall within *M. guttatus* populations. Elucidating the phylogenetic
397 relationships within the *M. guttatus* species complex has proven challenging (Beardsley *et al.*
398 2003; Beardsley *et al.* 2004), but the use of genome-wide sequences in a phylogenetic context
399 (Wagner *et al.* 2013) could provide a tool to establish the genetic relationships among
400 populations of this interfertile group.
401 The genetic structure observed in indigenous populations, allowed us to determine with
402 confidence that introduced populations in the UK are most genetically similar to populations
403 from the northern end of the native distribution. In particular, our results indicate that UK
404 samples are most genetically similar to the coastal perennial population of TSG (Figs. 2 and 3).
405 However, without additional sampling, the exact source for introduced populations in the UK
406 remains unknown. The genetic similarity within UK populations, which fall within a single clade

(Fig. 2), suggests that this part of the introduced range has been established either via a single introduction event or, perhaps more likely, via multiple introductions from closely related native populations. In order to determine the origin and number of introductions of UK *M. guttatus* with more accuracy, it will be necessary to conduct further sampling in the Alaskan end of the distribution, since historical records point to this region as a potential source for the first specimens of this taxon in Europe (Sims 1812; Pennell 1935, p. 116).

Despite a 50% reduction in nucleotide diversity (π) in non-native *M. guttatus*, UK populations still harbour a significant amount of diversity among populations. In particular, the estimate of nucleotide diversity at synonymous sites in UK populations (π_{syn} = 0.0325) is above the median estimate of neutral nucleotide diversity for outcrossing flowering plants (π = 0.0148), and much higher than the estimate for selfing taxa (π = 0.0035) (Leffler *et al.* 2012). The diversity observed in UK populations suggests that the bottleneck experienced by *M. guttatus* during invasion was weak enough to permit the maintenance of a significant amount of nucleotide diversity. In general, introduced populations tend to show reduced genetic diversity compared to native ones, but phenomena including large initial propagule numbers, multiple introductions, and admixture can result in equal or higher levels of genetic variation (Dlugosch & Parker 2008). The existence of non-synonymous variation in UK *M. guttatus* ($\pi_{non\text{-}syn}$ = 0.0035) is potentially important for naturalisation and spread in the new range, as introduced populations may be able to respond to new selective pressures from standing genetic variation, and are not necessarily limited by mutation rate (Prentis *et al.* 2008). Adaptation from standing

427 genetic variation is particularly important for invasive species as it may allow responding more
428 rapidly to novel selective pressures (Barrett & Schluter 2008).

429 Detecting selection in non-equilibrium populations

430 Our study uncovered several areas of the genome that bear the signature of positive selection.
431 The signal left by selective sweeps include increased frequency of previously rare derived
432 alleles, reduction of variation in linked sites, and increase in local linkage disequilibrium (Smith
433 & Haigh 1974; Barton 1998). However, past demographic changes can confound the pattern of a
434 selective sweep (Barton 1998; Barton & Etheridge 2004). This is particularly the case in
435 populations that have experienced a bottleneck and subsequent population growth with
436 recombination (Barton & Etheridge 2004; Nielsen *et al.* 2005). Furthermore, recombination
437 between lineages that either escape or not the population bottleneck will result in a
438 combination of short and long coalescent branches creating a sweep like pattern in the SFS
439 (Pavlidis *et al.* 2010). Thus, inferences of selection based on genome scans in non-equilibrium
440 populations must be done with caution.

441 The implementation that we used to detect the signature of positive selection
442 (*SweepFinder*), is relatively robust to the effects of bottlenecks and population expansion, and is
443 particularly powerful to detect recent selective sweeps (Nielsen *et al.* 2005; Pavlidis *et al.* 2010).
444 Recent studies have shown that *SweepFinder* performs better than other tests, such as the ω-
445 statistic, in non-equilibrium populations (Pavlidis *et al.* 2010). We consider that the very recent
446 introduction of *M. guttatus* into the UK, relatively weak population bottleneck, and lack of
447 within-UK population structure, make *SweepFinder* an appropriate method for detecting recent

448 and strong sweeps in the introduced range. Moreover, the independently analysed NA data
449 using the exact same SNP sampling strategy allowed us to mask shared sweeps and eliminate
450 false positives due solely to SNP sampling and variable site density. Therefore, we consider that
451 the genomic regions identified here are strong candidates for selective sweeps in the lineage
452 leading to UK populations. Nevertheless, it is important to recognise that we cannot currently
453 determine whether the selection events happened in an ancestral (unsampled) native
454 population or after the dispersal of *M. guttatus* to the UK. Establishing the timing of the
455 potential selective event is essential to determine if introduced populations can exploit novel
456 environments through (pre-)adaptations brought in from the native range, or whether adaptive
457 evolution occurs subsequent to dispersal during the establishment and spread phases of a
458 biological invasion (Maron *et al.* 2004; Colautti *et al.* 2009).

459 ## Genes within swept regions

460 The potential selective sweeps we detected in five *M. guttatus* chromosomes include
461 approximately 300 genes. As predicted for selective sweeps, we found reduced diversity in both
462 coding and non-coding genic regions under these sweeps (Fig S6). While it is possible to identify
463 candidate regions for selective sweeps and to determine the genes located under these sweeps,
464 it is not possible to know without direct testing which genes were the actual targets of
465 selection. However, the selective sweeps identified here contain genes involved in flowering
466 time, nutrient stress, and biotic stress (Table S2), and could be involved in adaptation to
467 different day lengths, soil types, novel pathogens, and general responses to stress (e.g., Hodgins
468 *et al.* 2012).

Of particular interest is the identification of selective sweeps in linkage group 8 (LG8). The selective sweeps at positions 2 Mbp (two million base pairs) and 5 Mbp of LG8 are exceptionally wide, consistent with very strong selection pulling a large haplotype block to high frequency very quickly giving very little time for recombination to break up linked sites. These regions of LG8 are gene rich suggesting that the width of the peak is not an artefact of reduced recombination associated with repetitive regions (Hellsten *et al.* 2013). Instead, these selective sweeps are located near or at a known inversion region of approximately 6Mb in length (Oneal *et al.* 2014). This inversion (*DIV1*), is polymorphic in the native range, and is associated with a number of morphological and life-history differences between annual and perennial ecotypes (Lowry and Willis 2010). Therefore this region is a good candidate for bearing adaptive variation that could be selected for in the lineage leading to the introduced populations. Mapping experiments in *M. guttatus* have demonstrated that a large region on LG8 is involved in critical photoperiod to flower (Friedman & Willis 2013). The selective sweep we identified on LG8 is located near (within 200,000 base pairs) one of the major QTLs for critical photoperiod identified in Friedman and Willis' study. Flowering time is a crucial component of fitness in seasonal environments, and therefore it is expected to be under selection in the introduced range. Studies of invasive species have often demonstrated the potential for rapid evolution of phenology in the introduced range (Maron *et al.* 2004; Colautti *et al.* 2009). The geographic distribution of introduced populations at high latitudes (approximately between 50° N and 60° N in the UK), suggests that these populations experience day length conditions similar to the northern end of the distribution of *M. guttatus* in North America. It would be of interest to

490 determine if genes involved in the control of flowering under long days are under selection in
491 northern indigenous or in introduced populations. One such candidate is *Migut.D02071*, a
492 phytochrome-associated protein phosphatase gene located in one of the selective sweep
493 regions in LG4 (Table S2). In *Arabidopsis thaliana*, a similar gene (*ATFYPP3*) participates in the
494 regulation of flowering time in long days (www.string-db.org). Preliminary data suggests that UK
495 populations require long days (16hr) to flower (Vallejo-Marín, unpublished). Future studies
496 could address whether flowering time has indeed played a role in the establishment of
497 introduced populations in high latitudes.

498 To conclude, our results demonstrate the enormous potential for whole-genome
499 sequencing studies to contribute to the study of non-native populations. With a modest
500 sequencing effort we were able to quickly obtain previously unavailable information on the
501 origin and diversity of introduced populations of *M. guttatus*, as well as on the genomic
502 consequences of biological introductions, including identifying potential regions under
503 selection. As whole genome reference data becomes available for other non-model organisms,
504 resequencing studies are likely to be increasingly used to study the history and consequences of
505 biological invasions, and to establish the contribution of adaptive processes to shaping the
506 genomes of rapidly evolving populations.

507 ## ACKNOWLEDGMENTS

508 We thank J. H. Willis for his support and advice, the *Mimulus* community for providing
509 information and sharing unpublished data, and to A. Twyford, J. Friedman, V. Sork and three

anonymous reviewers for comments on a previous version of the manuscript. This work was supported by a US National Science Foundation – National Plant Genome Initiative postdoctoral fellowship to JP (NSF, 1202778), and a UK Natural Environment Research Council grant (NERC, NE/J012645/1) and a Carnegie Trust Travel grant to MVM.

REFERENCES

Baker HG, Stebbins GL, eds (1965) *The Genetics of Colonizing Species* Academic Press, Silomar.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution,* **23**, 38-44.

Barton NH (1998) The effect of hitch-hiking on neutral genealogies. *Genetical Research,* **72**, 123-133.

Barton NH, Etheridge AM (2004) The effect of selection on genealogies. *Genetics,* **166**, 1115-1131.

Beardsley PM, Schoenig SE, Whittall JB, Olmstead RG (2004) Patterns of evolution in Western North American *Mimulus* (Phrymaceae). *American Journal of Botany,* **91**, 474-489.

Beardsley PM, Yen A, Olmstead RG (2003) AFLP phylogeny of *Mimulus* section *Erythranthe* and the evolution of hummingbird pollination. *Evolution,* **57**, 1397-1410.

526 Brandvain Y, Kenney AM, Flagel L, Coop G, Sweigart AL (2014) Speciation and introgression
527 between *Mimulus nasutus* and *Mimulus guttatus*. *PLOS* Genetics, **10**, e1004410.

528 Colautti RI, Maron JL, Barrett SCH (2009) Common garden comparisons of native and introduced
529 plant populations: latitudinal clines can obscure evolutionary inferences. *Evolutionary*
530 *Applications,* **2**, 187-199.

531 Cox GW (2004) *Alien Species and Evolution* Island Press, Washington.

532 Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide
533 genetic marker discovery and genotyping using next-generation sequencing. *Nature*
534 *Reviews Genetics,* **12**, 499-510.

535 DePristo MA, Banks E, Poplin R, et al (2011) A framework for variation discovery and genotyping
536 using next-generation DNA sequencing data. *Nature Genetics,* **43**, 491-498.

537 Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive
538 evolution, and the role of multiple introductions. *Molecular Ecology,* **17**, 431-449.

539 Dlugosch KM, Lai Z, Bonin A, Hierro J, Rieseberg LH (2013) Allele identification for
540 transcriptome-based population genomics in the invasive plant *Centaurea solstitialis*. *G3-*
541 *Genes Genomes Genetics,* **3**, 359-367.

542 Dole JA (1992) Reproductive assurance mechanisms in 3 taxa of the *Mimulus guttatus* complex
543 (Schrophulariaceae). *American Journal of Botany,* **79**, 650-659.

544 Ellegren H (2014) Genome sequencing and population genomics in non-model organisms.
545 *Trends in Ecology & Evolution,* **29**, 51-63.

546 Fishman L, Kelly AJ, Willis JH (2002) Minor quantitative trait loci underlie floral traits associated
547 with mating system divergence in *Mimulus*. *Evolution,* **56**, 2138-2155.

548 Friedman J, Willis JH (2013) Major QTLs for critical photoperiod and vernalization underlie
549 extensive variation in flowering in the *Mimulus guttatus* species complex. *New Phytologist,*
550 **199**, 571-583.

551 Gardeners' Chronicle (1852) *The Gardeners' Chronicle and Agricultural Gazette* Bradbury and
552 Evans, London.

553 Garvin MR, Saitoh K, Gharrett AJ (2010) Application of single nucleotide polymorphisms to non-
554 model species: a technical review. *Molecular Ecology Resources,* **10**, 915-934.

555 Grant AL (1924) A monograph of the genus *Mimulus*. *Annals of the Missouri Botanical Garden,*
556 **11**, 99-380.

557 Hall MC, Willis JH (2006) Divergent selection on flowering time contributes to local adaptation in
558 *Mimulus guttatus* populations. *Evolution,* **60**, 2466-2477.

559 Hellsten U, Wright KM, Jenkins J, et al (2013) Fine-scale variation in meiotic recombination in
560 *Mimulus* inferred from population shotgun sequencing. *Proceedings of the National*
561 *Academy of Sciences of the United States of America,* **110**, 19478-19482.

562 Helyar SJ, Hemmer-Hansen J, Bekkevold D, et al (2011) Application of SNPs for population
563 genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology*
564 *Resources,* **11**, 123-136.

565 Hodgins KA, Lai Z, Kurkowski K, Huang J, Rieseberg LH (2012) The molecular basis of
566 invasiveness: differences in gene expression of native and introduced common ragweed
567 (*Ambrosia artemisiifolia*) in stressful and benign environments. *Molecular Ecology*, **22**,
568 2496-2510.

569 Hulme PE (2009) Trade, transport and trouble: managing invasive species pathways in an era of
570 globalization. *Journal of Applied Ecology,* **46**, 10-18.

571 Hutter S, Vilella AJ, Rozas J (2006) Genome-wide DNA polymorphism analyses using *VariScan*.
572 *BMC bioinformatics,* **7**, 409.

573 Jahodová Š, Trybush S, Pyšek P, Wade M, Karp A (2007) Invasive species of *Heracleum* in
574 Europe: an insight into genetic relationships and invasion history. *Diversity and*
575 *Distributions,* **13**, 99-114.

576 Jombart T, Ahmed I (2011) adegenet 1.3-1: New tools for the analysis of genome-wide SNP data.
577 *Bioinformatics,* **27**, 3070-3071.

578 Lachmuth S, Durka W, Schurr FM (2010) The making of a rapid plant invader: genetic diversity
579 and differentiation in the native and invaded range of *Senecio inaequidens*. *Molecular*
580 *Ecology,* **19**, 3952-3967.

581 Langmead B, Salzberg SL (2014) Bowtie 2.2.1. Ultrafast and memory-efficient alignment of short
582 DNA sequences to the human genome. *Nature Methods,* **9**, 357-359.

583 Lee CE (2002) Evolutionary genetics of invasive species. *Trends in Ecology & Evolution,* **17**, 386-
584 391.

585 Leffler EM, Bullaughey K, Matute DR, Meyer, WK, Ségurel L, Venkat A, Andolfatto P, Przeworski
586 M (2012) Revisiting and old riddle: What determines genetic diversity levels within species?
587 *PLOS Biology*, **10**, e1001388.

588 Liu H, Pemberton RW (2009) Solitary invasive orchid bee outperforms co-occurring native bees
589 to promote fruit set of an invasive *Solanum*. *Oecologia,* **159**, 515-525.

590 Lowry DB, Rockwood RC, Willis JH (2008) Ecological reproductive isolation of coast and inland
591 races of *Mimulus guttatus*. *Evolution,* **62**, 2196-2214.

592 Lowry DB, Hall MC, Salt DE, Willis JH (2009) Genetic and physiological basis of adaptive salt
593 tolerance divergence between coastal and inland *Mimulus guttatus*. *New Phytologist,* **183**,
594 776-788.

595 Lowry DB, Rockwood RC, Willis JH (2008) Ecological reproductive isolation of coast and inland
596 races of *Mimulus guttatus*. *Evolution,* **62**, 2196-2214.

597 Lowry DB, Willis JH (2010) A widespread chromosomal inversion polymorphism contributes to a
598 major life-history transition, local adaptation, and reproductive isolation. *PLoS Biology,* **8**,
599 e1000500.

600 Macnair MR, Watkins AD (1983) The fitness of the copper tolerance gene of *Mimulus guttatus* in
601 uncontaminated soil. *New Phytologist,* **95**, 133-137.

602 Maron JL, Vila M, Bommarco R, Elmendorf S, Beardsley P (2004) Rapid evolution of an invasive
603 plant. *Ecological Monographs,* **74**, 261-280.

604 Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective sweeps.
605 *Trends in Ecology & Evolution,* **28**, 659-669.

606 Milne RI, Abbott RJ (2000) Origin and evolution of invasive naturalized material of
607 *Rhododendron ponticum* L. in the British Isles. *Molecular Ecology,* **9**, 541-556.

608 Murren CJ, Chang CC, Dudash MR (2009) Patterns of selection of two North American native and
609 nonnative populations of monkeyflower (Phrymaceae). *New Phytologist,* **183**, 691-701.

610 Nesom G (2012) Taxonomy of *Erythranthe* Sect. *Simiola* (Phrymaceae) in the USA and Mexico.
611 *Phytoneuron,* **40**, 1-123.

612 Ness RW, Siol M, Barrett SCH (2012) Genomic consequences of transitions from cross- to self-
613 fertilization on the efficacy of selection in three independently derived selfing plants. *BMC*
614 *Genomics,* **13**, 611.

615 Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C (2005) Genomic scans for
616 selective sweeps using SNP data. *Genome Research,* **15**, 1566-1575.

617 Oneal E, Lowry DB, Wright KM, Zhu Z, Willis JH (2014) Divergent population structure and
618 climate associations of a chromosomal inversion polymorphism across the *Mimulus*
619 *guttatus* species complex. *Molecular Ecology*, **23**, 2844-2860.

620 Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R
621 language. *Bioinformatics,* **20**, 289-290.

622 Parker PF (1975) *Mimulus* in Great Britain: A cytotaxonomic note. *New Phytologist,* **74**, 155-160.

623 Pavlidis P, Jensen JD, Stephan W (2010) Searching for footprints of positive selection in whole-
624 genome SNP data from nonequilibrium populations. *Genetics,* **185**, 907-922.

625 Pennell FW (1935) The Scrophulariaceae of eastern temperate North America. *The Academy of*
626 *Natural Sciences of Philadelphia,* **1**, 1-650.

627 Pennell F (1951) *Mimulus*. In: *Illustrated Flora of the Pacific States* (eds. Abrams LR, Ferris RS),
628 vol. 3 edn, pp. 688-731 Stanford University Press.

629 Phillips BL, Shine R (2006) An invasive species induces rapid adaptive change in a native
630 predator: cane toads and black snakes in Australia. *Proceedings of the Royal Society B-*
631 *Biological Sciences,* **273**, 1545-1550.

632 Pimentel D, ed (2002) *Biological Invasions: Economic and Environmental Costs of Alien Plant,*
633 *Animal, and Microbe Species* CRC Press, Boca Raton, Florida.

634 Prentis PJ, Wilson JRU, Dormontt EE, Richardson DM, Lowe AJ (2008) Adaptive evolution in
635 invasive species. *Trends in Plant Science,* **13**, 288-294.

636 Prentis PJ, Woolfit M, Thomas-Hall SR, et al (2010) Massively parallel sequencing and analysis of
637 expressed sequence tags in a successful invasive plant. *Annals of Botany,* **106**, 1009-1017.

638 Preston CD, Pearman DA, Dines TD, eds (2002) *New Atlas of the British and Irish Flora* Oxford
639 University Press, Oxford.

640 R Development Core Team (2014) R. A language and environment for statistical computing. R
641 Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. **ver.**
642 **3.0.3**.

643 Rambaut A (2014) *FigTree* version 1.4.0. *http://tree.bio.ed.ac.uk/software/figtree/[accessed*
644 *March 12, 2014]*.

645 Ricciardi A, Hoopes MF, Marchetti MP, Lockwood JL (2013) Progress toward understanding the
646 ecological impacts of nonnative species. *Ecological Monographs,* **83**, 263-282.

647 Ritland K (1990) Inferences about inbreeding depression based on changes of the inbreeding
648 coefficient. *Evolution,* **44**, 1230-1241.

649 Roberts RH (1964) *Mimulus* hybrids in Britain. *Watsonia,* **6**, 70-75.

650 Roda F, Ambrose L, Walter GM, et al (2013) Genomic evidence for the parallel evolution of
651 coastal forms in the *Senecio lautus* complex. *Molecular Ecology,* **22**, 2941-2952.

652 Savolainen O, Lascoux M, Merila J (2013) Ecological genomics of local adaptation. *Nature*
653 *Reviews Genetics,* **14**, 807-820.

654 Sax DF, Stachowicz JJ, Brown JH, et al (2007) Ecological and evolutionary insights from species
655 invasions. *Trends in Ecology & Evolution,* **22**, 465-471.

656 Silverside AJ (1990) A new binomial in *Mimulus*. *Watsonia,* **18**, 210-212.

657 Sims J (1812) *Curtis's Botanical Magazine. Flower-Garden Displayed. Vol. 35.* Sherwood, Neeley
658 & Jones, London.

659 Siol M, Wright SI, Barrett SCH (2010) The population genomics of plant adaptation. *New*
660 *Phytologist,* **188**, 313-332.

661 Smith JM, Haigh J (1974) The hitchhiking effect of a favorable gene. *Genetical Research,* **23**, 23-
662 35.

663 Stace C (2010) *New Flora of the British Isles*, Third Edition edn. Cambridge University Press,
664 Cambridge.

665 Sweigart AL, Martin NH, Willis JH (2008) Patterns of nucleotide variation and reproductive
666 isolation between a *Mimulus* allotetraploid and its progenitor species. *Molecular Ecology,*
667 **17**, 2089-2100.

668 Tarnowska K, Daguin-Thiebaut C, Pain-Devin S, Viard F (2013) Nuclear and mitochondrial genetic
669 variability of an old invader, *Dreissena polymorpha* (Bivalvia), in French river basins.
670 *Biological Invasions,* **15**, 2547-2561.

671 Tokarska-Guzik B, Dajdok Z (2010) NOBANIS. Invasive Alien Species Fact Sheet: *Mimulus*
672 *guttatus*. Online Database of the European Network on Invasive Alien Species – NOBANIS
673 www.nobanis.org, Date of access 1/03/2014.

674 Truscott AM, Soulsby C, Palmer SCF, Newell L, Hulme PE (2006) The dispersal characteristics of
675 the invasive plant *Mimulus guttatus* and the ecological significance of increased occurrence
676 of high-flow events. *Journal of Ecology,* **94**, 1080-1091.

677 Twyford AD, Ennos RA (2012) Next-generation sequencing as a tool for plant ecology and
678 evolution. *Plant Ecology & Diversity,* **5**, 411-413.

679 Vallejo-Marín M, Lye GC (2013) Hybridisation and genetic diversity in introduced *Mimulus*
680 (Phrymaceae). *Heredity,* **110**, 111-122.

van Kleunen M (2007) Adaptive genetic differentiation in life-history traits between populations of *Mimulus guttatus* with annual and perennial life-cycles. *Evolutionary Ecology,* **21**, 185-199.

van Kleunen M, Fischer M (2008) Adaptive rather than non-adaptive evolution of *Mimulus guttatus* in its invasive range. *Basic and Applied Ecology,* **9**, 213-223.

Vickery RK (1959) Barriers to gene exchange within *Mimulus guttatus* (Scrophulariaceae). *Evolution,* **13**, 300-310.

Vila M, Espinar JL, Hejda M, et al (2011) Ecological impacts of invasive alien plants: a meta-analysis of their effects on species, communities and ecosystems. *Ecology Letters,* **14**, 702-708.

Vos P, Hogers R, Bleeker M, et al (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic acids research,* **23**, 4407-4414.

Wagner CE, Keller I, Wittwer S, et al (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology,* **22**, 787-798.

Williams F, Eschen R, Harris A, et al (2010) The economic cost of invasive non-native species on Great Britain. *CABI report, 198pp*.

698 Wu CA, Lowry DB, Cooley AM, Wright KM, Lee YW, Willis JH (2007) *Mimulus* is an emerging
699 model system for the integration of ecological and genomic studies. *Heredity,* **100**, 220-230.

700 Zhang Z, Li J, Zhao X, Wang J, Wong GK, Yu J (2006) KaKs_calculator: Calculating *Ka* and *Ks*
701 through model selection and model averaging. *Genomics Proteomics & Bioinformatics,* **4**,
702 259-263.

Data Accessibility

DNA sequences: NCBI SRA (individual population codes in parentheses): SRR1462346 (10-AYR-10), SRR1475232 (10-CER-10), SRR1475385 (10-DBL-20), SRR1481643 (10-HOU-17), SRR1481644 (10-QUA-47), SRR1482404 (10-TOM-23), SRR1482405 (12-CRO-5), SRR1482406 (12-PAC-39), SRR1482407 (12-VIC-18), SRR1482409 (12-TRE-17).

SNP Genotype data: Dryad doi:10.5061/dryad.3gp32

Author Contributions

MVM and JP designed the research. MVM made the field surveys and collected the samples. JP and MVM organized laboratory work. JP generated the genomic data and conducted the bioinformatic analyses. JP and MVM analysed the genomic data. MVM and JP wrote the manuscript.

TABLES

Table 1. Populations sampled for genome resequencing of *Mimulus guttatus* and related taxa. A single individual was sequenced per population, except for IM, where sequence data was available for nine individuals. Life history and, for native populations, their classification as coastal or inland, is provided for each population when available (Lowry *et al.* 2008; Lowry & Willis 2010).

Population	Taxon	Latitude	Longitude	Life history	Coastal/Inland	Location
	***Mimulus guttatus* (Native)**					
PED		32.711	-110.628	Perennial	Inland	San Pedro River, Pinal Co., AZ
MED		37.829	-120.345	Annual	Inland	Moccasin, Tuolumne Co., CA
REM		38.859	-122.410	--	Inland	Rumsey, Yolo Co., CA
LMC		38.864	-123.084	Annual	Inland	Yorkville, Mendocino Co., CA
SWB		39.036	-123.691	Perennial	Coastal	Sperm Whale Beach, Mendocino Co., CA
BOG		41.924	-118.804	Perennial	Inland	Bog Hot Springs, Humboldt, Co., NV
MAR		43.479	-123.294	Annual	Inland	Marshanne Landing , Douglas Co., OR
DUN		43.893	-124.130	Perennial	Coastal	Dunes, Lane Co., OR
IM		44.401	-122.151	Annual	Inland	Iron Mountain, Linn Co., OR
AHQ		44.431	-110.813	Perennial	Inland	Lonestar Basin Thermal Spring, Teton Co., WY
YJS		44.951	-114.585	Perennial	Inland	Yellowjacket creek, Lemhi Co., ID
TSG		53.419	-131.916	Perennial	Coastal	Graham Island, Haida Gwaii (Queen Charlotte Islands), British Columbia, Canada
	***Mimulus guttatus* (Introduced)**					
CRO		50.163	-5.293	Perennial	--	Crowan, Cornwall
TRE		50.498	-4.465	Perennial	--	Tremar Coombe, Cornwall, England
HOU		51.097	-1.508	Perennial	--	Houghton Lodge, Hampshire, England

CER		53.006	-3.549	Perennial	--	Cerrigydrudion, Denbigshire, Wales
VIC		54.763	-7.454	Perennial	--	Victoria Bridge, Northern Ireland
AYR		55.461	-4.625	Perennial	--	Ayr, Ayrshire, Scotland
DBL		56.197	-3.965	Perennial	--	Dunblane, Perthshire, Scotland
TOM		57.255	-3.368	Perennial	--	Tomintoul, Moray, Scotland
PAC		57.355	-3.336	Perennial	--	Packhorse Bridge, Speyside, Scotland
QUA		60.105	-1.227	Perennial	--	Quarff, Shetland Islands
	Outgroups					
SF	*M. nasutus*	45.635	-120.914	Annual	Inland	Sherars Falls, Wasco Co., OR
MCN	*M. cupriphilus*	37.912	-120.724	Annual	Inland	McNulty Mine, Calaveras, Co., CA
CVP	*M. platycalyx*	38.372	-123.055	Annual	Inland/Coastal	Coleman Valley Road, Sonoma Co., CA
EBR	*M. micranthus*	39.631	-123.532	Annual	Inland	Branscomb, Mendocino Co., CA
DENT	*M. dentilobus*	NA	NA	--	--	NA

717

FIGURES

Figure 1. Location of the 22 *Mimulus guttatus* populations sampled in the native range in North America (left-hand side panel), and in the introduced range the United Kingdom (right-hand side panel). Notice the different scales in the two maps. The colour of the symbols corresponds to the clades shown in the neighbour joining tree in Fig. 2.

Figure 2. Relationships between native (North America) and introduced (United Kingdom) populations of *Mimulus guttatus* inferred from 1,400,000 SNPs, sampled at a density of 100,000 SNPs per chromosome. The neighbour joining tree was built from a matrix of pairwise genetic distance (*p*-distance) of 30 individuals of *M. guttatus,* four related taxa, and rooted with *M. dentilobus*. Branches leading to introduced populations are shown in purple. The two clades containing native *M. guttatus* in the North and South groups (see Results) are shown in blue and red, respectively. All nodes at the population level have a 100/100 bootstrap support. Population names as in Table 1.

Figure 3. Principal component analysis of twelve native (North America) and ten introduced (United Kingdom) populations of *Mimulus guttatus.* Introduced populations cluster in the lower left quadrat. Only one individual of the Iron Mountain (IM) population was included in this analysis. Colours denote values at first two principal component (PC) axes. Population names as in Table 1.

Figure 4. Evidence of hard selective sweeps within UK *M. guttatus* was found on 5 of the 14 linkage groups using *SweepFinder* (Nielsen *et al.* 2005). The figure shows selective sweeps after

738 masking for shared sweeps between native and introduced populations (see Methods). Dashed
739 line indicates genome-wide 1% outlier cutoff based on the composite likelihood ratio (CLR)
740 statistic. Only linkage groups with sweeps exceeding 1% outlier cutoff are shown here; CLR
741 profiles for all linkage groups are shown for both native and introduced populations in the
742 supplementary materials.

SUPPLEMENTARY MATERIAL

Figure S1. Distribution of genome-wide pairwise nucleotide diversity (i.e., the average number of nucleotide differences per site between two sequences, π) in native and introduced populations of the yellow monkeyflower, *Mimulus guttatus*. The left-hand side panel shows nucleotide diversity of 10 introduced populations in the United Kingdom, and the right-hand side panel shows nucleotide diversity for 10 native populations in North America (same 10 used in sweep analysis). Nucleotide diversity was calculated using windows of 50,000 genotyped base pairs (bp) and 1,000 bp overlapping steps.

Figure S2. Genome-wide pattern of pairwise nucleotide diversity (π) in 10 introduced populations of *Mimulus guttatus* in the United Kingdom. Each row represents one of the 14 major linkage groups (scaffolds). Nucleotide diversity was calculated using windows of 50,000 genotyped base pairs (bp) and 1,000 bp overlapping steps. The x-axis indicates the midpoint position of each window with respect to the *M. guttatus v 2.0* reference genome (www.phytozome.org).

Figure S3. Genome-wide of pairwise nucleotide diversity (π) in 10 native populations of *Mimulus guttatus* in North America (AHQ, BOG, DUN, IM, LMC, MAR, PED, SWB, TSG, YJS)**.** Each row represents one of the 14 major linkage groups (scaffolds). Nucleotide diversity was calculated using windows of 50,000 genotyped base pairs (bp) and 1,000 bp overlapping steps. The x-axis indicates the midpoint position of each window with respect to the *M. guttatus v 2.0* reference

762 genome (www.phytozome.org).

763 **Figure S4.** Distribution of non-synonymous ($\pi_{non\text{-}syn}$) and synonymous (π_{syn}) pairwise nucleotide
764 diversity in 10 introduced (left-hand side panels; United Kingdom), and in 10 native populations
765 (right-hand side panels; North America) of *Mimulus guttatus*.

766 **Figure S5.** Selective sweeps identified in introduced populations (United Kingdom) of the yellow
767 monkeyflower, *Mimulus guttatus*, across the 14 major linkage groups (scaffolds) in this species.
768 The composite likelihood ratio (CLR) measures departures in the site frequency spectrum from
769 background levels observed within linkage groups, and provide evidence of selective sweeps
770 (Nielsen *et al.* 2005). Black dots indicate CLR scores above the 1% outlier cutoff, after masking
771 (omitting) regions with significant CLR scores in North American samples. The x-axis indicates
772 the position in base pairs with respect to the *M. guttatus v 2.0* reference genome
773 (www.phytozome.org). For full details of the analysis see Methods section.

774 **Figure S6.** Evidence for reduced pairwise nucleotide diversity (π and π_{syn}) in genes located
775 within (sweep) *vs.* outside selective sweeps (no sweep) of introduced populations of *Mimulus*
776 *guttatus* in the United Kingdom. The left-hand side panel shows pairwise nucleotide diversity
777 across all sites within a gene (π), and the left-hand side panel shows pairwise nucleotide
778 diversity in synonymous sites only (π_{syn}). For the left panel, nucleotide diversity was calculated
779 using windows of 50,000 genotyped base pairs (bp) and 1,000 bp overlapping steps.
780 Synonymous site nucleotide diversity was calculated for all genes within swept region. Boxplots
781 show the median (horizontal line), 1st and 3rd quartiles (top and bottom of the box), and the 1.5

782 interquartile range (whiskers).

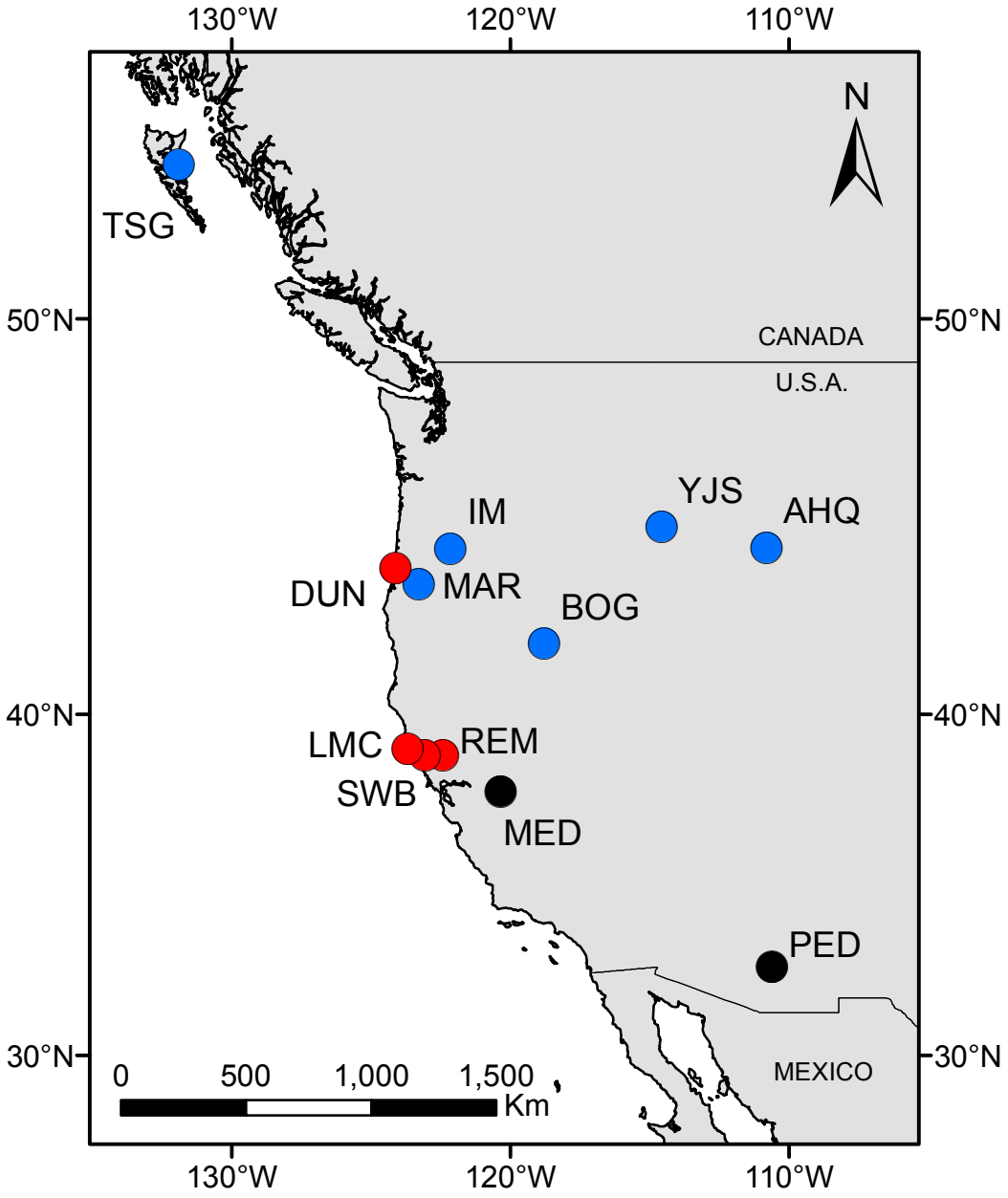
783 **Figure S7.** Genome-wide distribution of genotyped base pairs per individual across ten
784 individuals from ten introduced populations of *Mimulus guttatus* in the United Kingdom (UK).
785 The chart shows, for instance, that nine out of ten individuals were successfully genotyped at
786 more than 18 million base pairs.

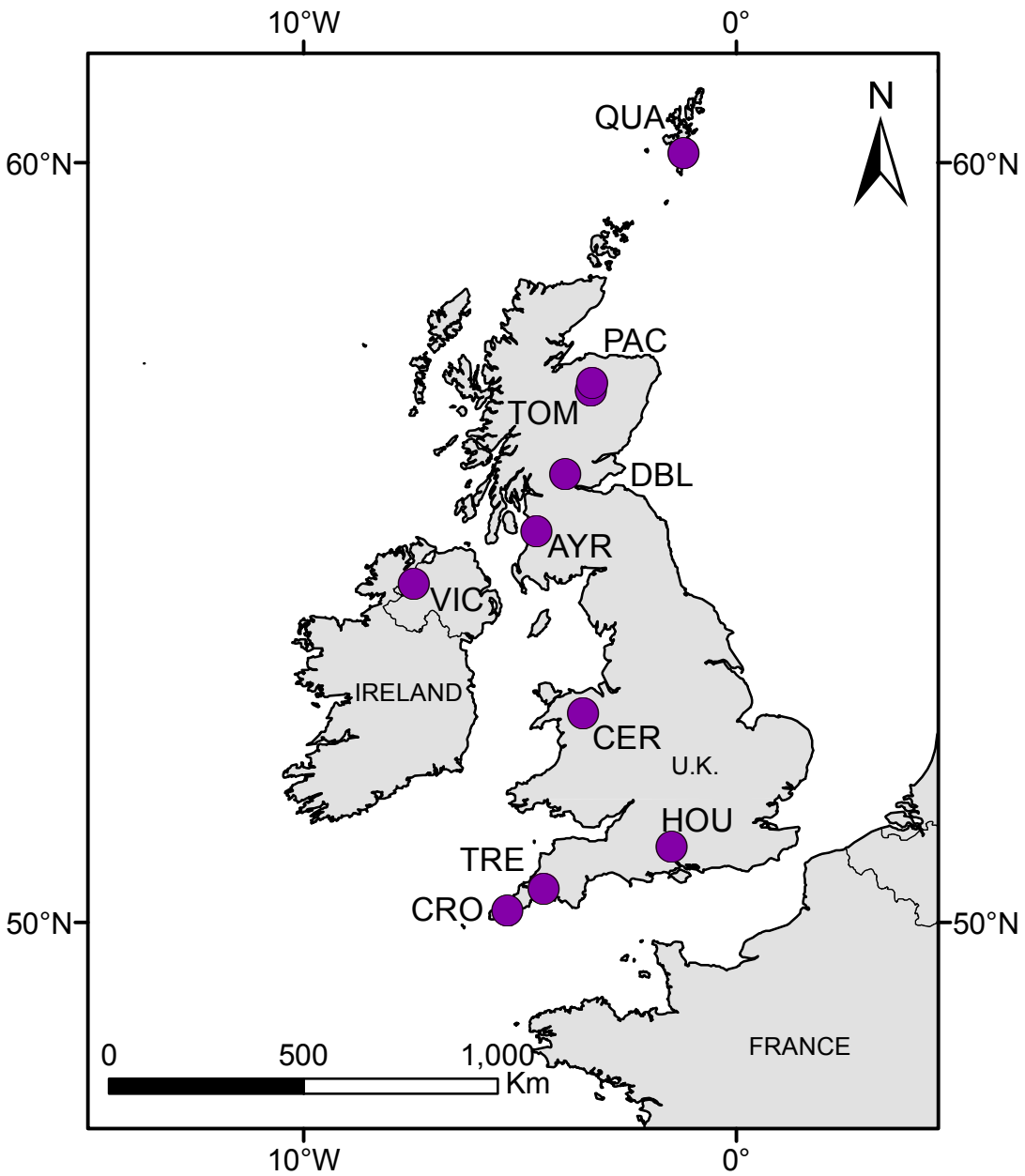
787 **Figure S8.** Principal component analysis (PCA) of 30 individuals of *Mimulus guttatus* from both
788 native and introduced ranges, including multiple individuals of the Iron Mountain (IM)
789 population. This PCA was based on 10,000 SNPs per linkage group for a total of 140,000 SNPs
790 across the genome. Introduced populations fall in the bottom left quadrat. Population codes are
791 defined in Table 1.

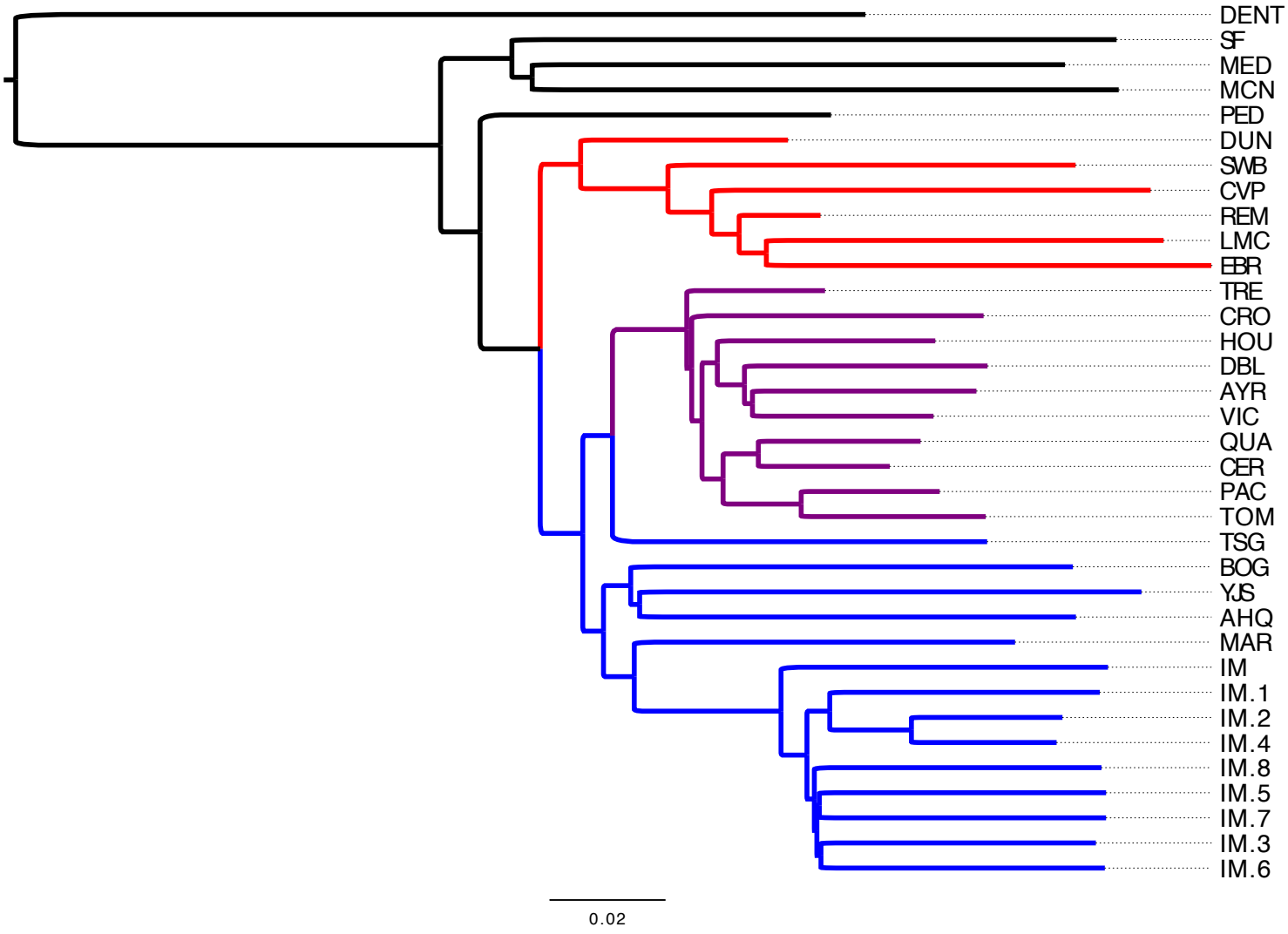
792 **Table S1.** Annotated list of 299 genes located within the selective sweeps identified in
793 introduced populations of the yellow monkeyflower, *Mimulus guttatus,* in the United Kingdom.
794 The location of each gene (start and end), and of the selective sweep that contains it, is given in
795 base pairs (bp) with respect to the *M. guttatus v. 2.0* reference genome (www.phytozome.org).

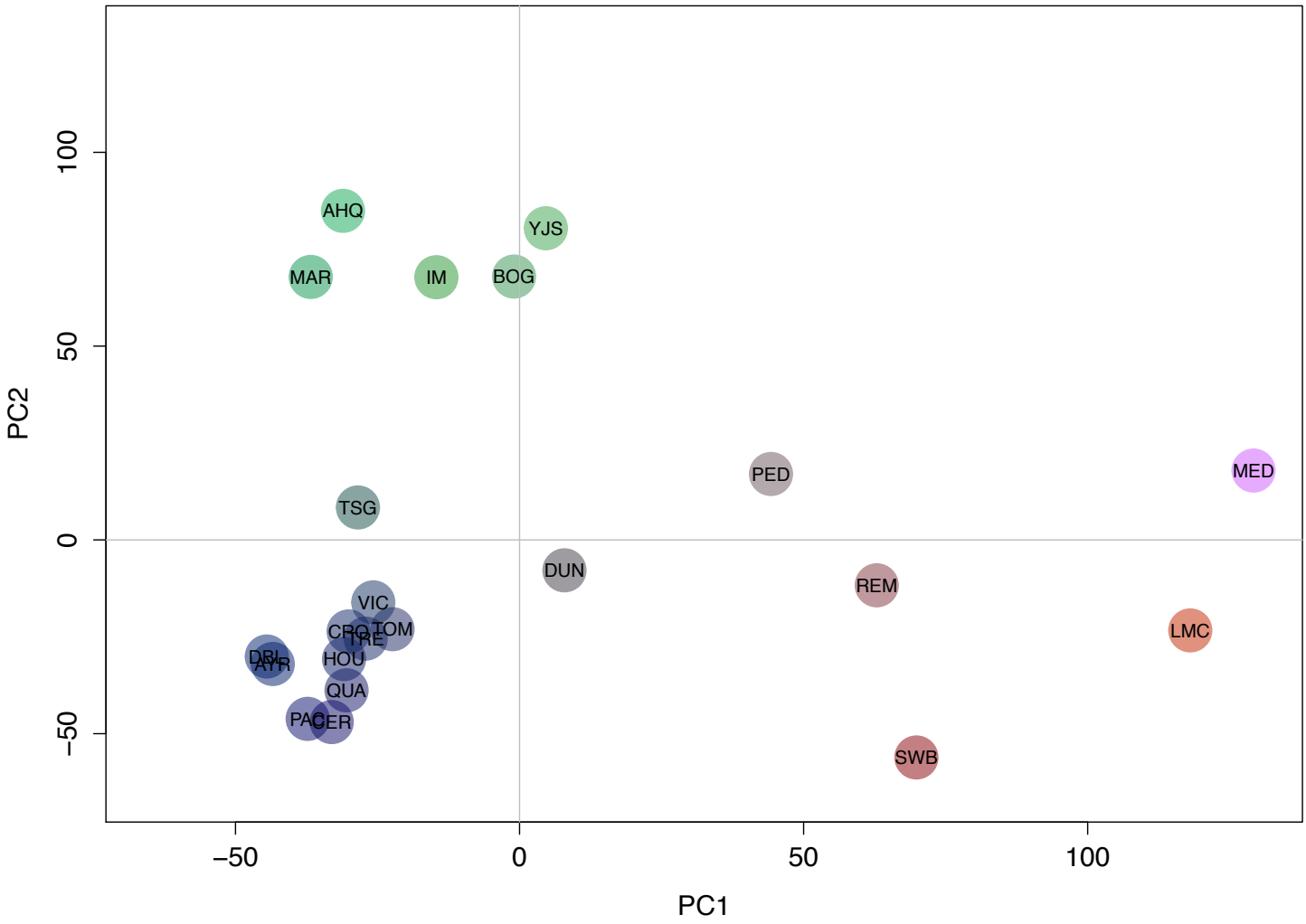
796 **Table S2.** Annotated list of a subset of 28 genes within the selective sweeps described in Table
797 S1 that display reduced synonymous diversity ($\pi_{syn} < 0.01$).

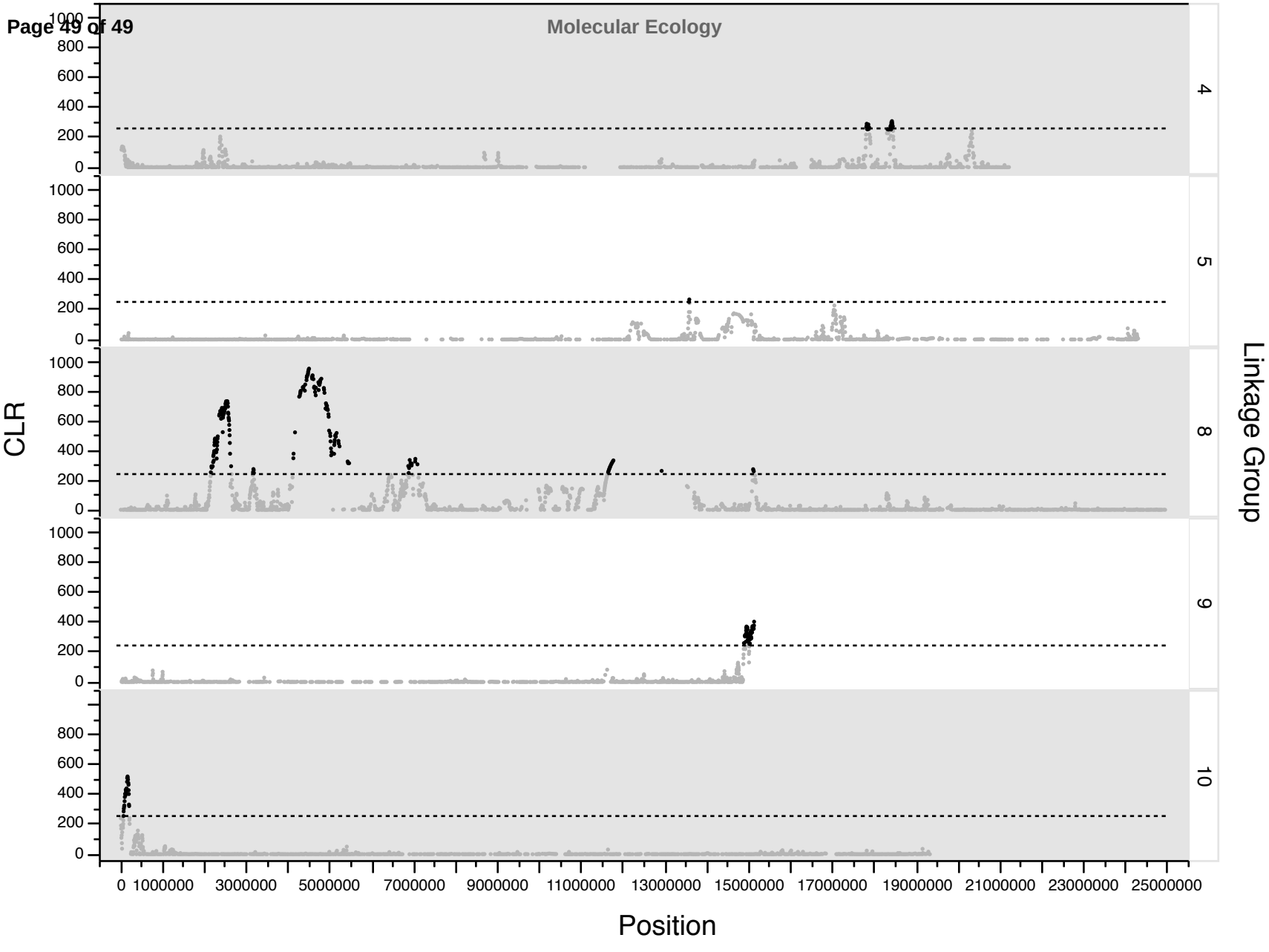
798 **Supplementary File: Bioinformatics.** Commands used for sequence analysis, quality control, and
799 genotyping.


130°W
120°W
110°W
N
TSG
50°N
CANADA
U.S.A.
YJS
AHQ
IM
DUN
MAR
BOG
40°N
LMC
REM
SWB
MED
PED
30°N
MEXICO
0
500
1,000
1,500
Km


10°W
0°
N
QUA
60°N
PAC
TOM
DBL
AYR
VIC
IRELAND
CER
U.K.
HOU
TRE
CRO
50°N
FRANCE
0
500
1,000
Km


DENT
SF
MED
MCN
PED
DUN
SWB
CVP
REM
LMC
EBR
TRE
CRO
HOU
DBL
AYR
VIC
QUA
CER
PAC
TOM
TSG
BOG
YJS
AHQ
MAR
IM
IM.1
IM.2
IM.4
IM.8
IM.5
IM.7
IM.3
IM.6
0.02


AHQ
YJS
MAR
IM
BOG
PED
MED
TSG
DUN
REM
VIC
CRO
TOM
TRE
LMC
DRI
AYR
HOU
QUA
PACER
SWB
100
50
0
−50
PC2
−50
0
50
100
PC1


CLR
Linkage Group
Position
4
5
8
9
10
0
200
400
600
800
1000
0 1000000
3000000
5000000
7000000
9000000
11000000
13000000
15000000
17000000
19000000
21000000
23000000
25000000